

# Diseño de una plantilla electrónica de evaluación de sedes web científicas para la creación de recursos de enseñanza-aprendizaje (alemán-inglés-español)<sup>1</sup>

Designing an evaluation template of scientific web sites for the implementation of teaching and learning materials (German-English-Spanish)

MIRIAM SEGHIRI<sup>2</sup>

[seghiri@uma.es](mailto:seghiri@uma.es)

Universidad de Málaga, España

## Resumen:

A día de hoy las herramientas de procesamiento de lenguaje natural (PLN) para la creación automática o semiautomática de herramientas de enseñanza-aprendizaje se basan, en su práctica totalidad, en corpus de textos descargados de la red Internet. Tal es el caso de la plataforma de enseñanza-aprendizaje trilingüe (alemán-

## Abstract:

Nowadays, most of the tools using natural language processing (NLP) for automatic or semi-automatic creation of teaching and learning materials are based on texts called corpora downloaded from the Internet. This is the case of the TELL-ME project that holds an e-learning platform for scientific and medical staff in three

1 El presente trabajo se enmarca en el seno del proyecto europeo Leonardo TELL-ME (Towards European Language Learning for MEDical professionals, ref. 517937-LLP-2011-UK-LEONARDO-LMP). Asimismo, se enmarca parcialmente en el proyecto europeo Marie Curie EXPERT (EXPloiting Empirical appRoaches to Translation, ref. 317471-FP7-PEOPLE-2012-ITN) y en los proyectos de I+D nacionales INTELITERM (Sistema inteligente de gestión terminológica para traductores, ref. FFI2012-38881), autonómicos TERMITUR (Diccionario inteligente TERMINológico para el sector TURístico, ref. HUM-2754) y locales TRADICOR (Sistema de gestión de corpus para la innovación, ref. PIE 13-054).

2 **Dirección para correspondencia (correspondence address):**

Miriam Seghiri. Dpto. de Traducción e Interpretación. Facultad de Filosofía y Letras. Universidad de Málaga. Campus de Teatinos, s/n. 29071 Málaga (España).

inglés- español) para personal científico y sanitario desarrollada en el proyecto europeo TELL-ME. A la hora de realizar búsquedas en la red con objeto de compilar un corpus, será imprescindible asegurar la fiabilidad de la muestra compilada para garantizar, a su vez, la calidad de los recursos que se implementen a partir de ella. Para ello es necesario la selección de unos parámetros e indicadores claros que tengan en cuenta las necesidades informativas del usuario (según el grado de especialidad), el tiempo del que se disponga para realizar dicho análisis así como el volumen de datos objeto de análisis. Una vez seleccionados, se puede proceder a la implementación de una plantilla electrónica para la evaluación de sedes web científicas en línea a partir de la cual se compilará el corpus para la creación de recursos de enseñanza-aprendizaje.

**Palabras clave:**

Evaluación; sedes web; ciencia; medicina; enseñanza-aprendizaje; parámetros; indicadores.

languages (German, English and Spanish). However, it is essential to ensure the reliability of the sample compiled to ensure, in turn, the quality of the resources that are implemented from it. A clear selection of parameters and indicators must be carried out. It is also important to take into account user needs (depending on the degree of specialization), the time available to perform the analysis, and finally, the volume of data being analyzed. Once these parameters and indicators have been selected, it will be possible to proceed to the implementation of an electronic template for the evaluation of scientific web sites online from which the corpus to create teaching-learning materials is compiled.

**Key words:**

Evaluation; web sites; science; medicine; teaching and learning; parameters; indicators.

**Résumé :**

Aujourd'hui, les outils de traitement automatique du langage naturel (TALN) pour la création automatique ou semi-automatique d'outils d'enseignement et d'apprentissage sont presque entièrement basés sur des corpus de textes téléchargés d'Internet. Tel est le cas de la plate-forme développée dans le projet européen TELL -ME pour l'enseignement et l'apprentissage des langues (allemand-anglais- espagnol) pour le personnel scientifique et médical. Cependant, il est essentiel d'assurer la fiabilité des textes électroniques compilés pour assurer, à son tour, la qualité des ressources qui sont développées à partir d'elle. Il faut sélectionner des paramètres et des indicateurs clairs qui tiennent compte des besoins d'information de l'utilisateur (selon le degré de spécialisation); du temps disponible pour effectuer l'analyse; et, du volume de données en cours d'analyse. Une fois sélectionnés, on va présenter un modèle électronique pour l'évaluation des sites web scientifiques en ligne à partir duquel est compilé le corpus pour la création de ressources d'enseignement et d'apprentissage.

**Mots clés:**

Évaluation; sites web; la science; la médecine; l'enseignement et l'apprentissage; paramètres; indicateurs.

Fecha de recepción: 15-4-2014

Fecha de aceptación: 16-6-2015

## 1. Introducción

Hoy en día, no cabe duda de que en toda Europa existe una creciente oleada de desplazamientos, muchos de ellos por trabajo, como puede ser el caso de profesionales sanitarios en busca de empleo fuera de su país de origen.

Así, por lo que se refiere a España, según Astorelli (2012):

Más de 300.000 españoles se han ido al extranjero desde el inicio de la crisis en 2008. [...] Entre los nuevos exiliados predominan los jóvenes de 25 a 45 años, con un alto perfil profesional y universitario. [Los puestos que más cubren estos jóvenes están] relacionados con la investigación, la medicina y la biología. [Y los destinos principales son] los países europeos desarrollados, como Alemania [...] o el Reino Unido [que] destacan por la demanda de personal sanitario para incorporar a sus hospitales.

A ello se suma el hecho de que España también es un destino preferente en el denominado *turismo sanitario*.

En este sentido, Rivera (2013) subraya que:

[e]l turismo sanitario genera 75.000 millones al año y mueve entre 30 y 50 millones de viajeros al año. [Así,] el sector privado busca turistas extranjeros para que se operen o sigan un tratamiento en España. Se ha constituido la plataforma Tourism & Health Spain, integrada por clínicas, hoteles y aseguradoras y más de 6.500 profesionales sanitarios.

Sin embargo, para que este flujo de personas pueda ofrecer o disfrutar de asistencia sanitaria será necesario hacer desaparecer la barrera del idioma. Con este objetivo en mente, el consorcio del proyecto TELL-ME<sup>3</sup>, pretende crear una plataforma de enseñanza-aprendizaje para mejorar la comunicación médica en tres lenguas europeas de gran demanda —el alemán, el inglés, y el español— y está pensado para ayudar a tres grupos de usuarios distintos, a saber, los profesionales médicos que desean aprender un idioma que les sirva para trabajar en el extranjero, profesio-

---

3 Para más información en torno al proyecto *TELL-ME (Towards European Language Learning for MEDical professionals)* consúltese la siguiente dirección URL: <http://www.tellme-project.eu/es>.

nales médicos que quieren aprender un idioma para poder comunicarse con los pacientes extranjeros que van a su consulta, y los pacientes cuya lengua materna es distinta a la del país que les presta asistencia médica. De esta forma, todos los usuarios que lo deseen pueden registrarse en la plataforma de enseñanza-aprendizaje (<https://clg2t.wlv.ac.uk/moodle-development/login/index.php>) creada para ofrecer cursos sobre discurso médico e interacción médico-paciente en español, inglés y alemán. La plataforma pone a disposición de los usuarios una gran multitud de recursos de enseñanza-aprendizaje, entre los que destacan diccionarios de imágenes, actividades de autoevaluación o glosarios de términos médicos, todos ellos basados en corpus (cfr. Figura 1).

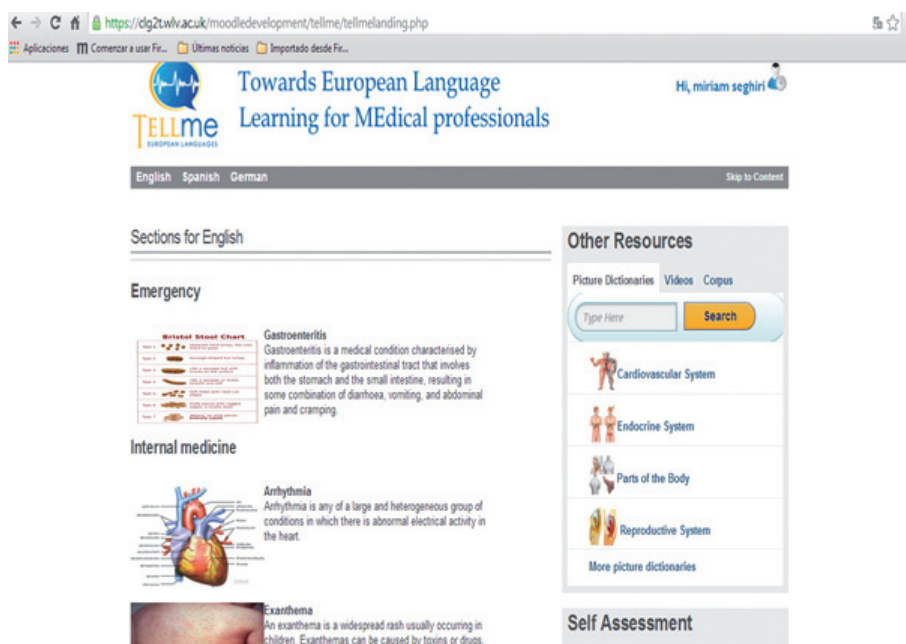


Figura 1. Página de inicio de la plataforma de enseñanza-aprendizaje TELL-ME.

Sin embargo, será necesario asegurar la calidad de los textos que integran el corpus compilado mediante unos parámetros claros y fiables para que las herramientas creadas a partir de él en la plataforma TELL-ME también lo sean.

## 2. Objetivos

La red Internet se ha convertido en la gran Biblioteca de Alejandría del siglo xxi. Sin embargo, su uso no está exento de peligros pues, tal y como subraya Austermühl (2001, p. 52), “[f]inding data on the web is not problem at all. But finding reliable information is rather a difficult task. And finding the information you really need can be very time-consuming and often frustrating”. En este sentido, no podemos olvidar que para compilar un corpus virtual no es tan importante la cantidad de los documentos compilados como su *calidad*, que es crucial para la muestra compilada (cfr. Seghiri, 2011, pp. 13-30). Atendiendo a las necesidades del proyecto, a la Universidad de Málaga se encargó de compilar un corpus de texto, virtual, trilingüe (alemán-español-inglés) que fuera fiable y de calidad, pues a partir de él se desarrollarán los contenidos de la plataforma de enseñanza-aprendizaje. Con esta idea en mente, los objetivos principales del presente trabajo, y que desarrollaremos a continuación, son los siguientes:

- a) Establecer unos criterios claros para la evaluación de recursos electrónicos que tengan en cuenta *qué* se evalúa, en función de una jerarquía conceptual, donde nos centraremos en los sitios o sedes web.
- b) Realizar un excursio por los principales parámetros e indicadores que permiten al usuario saber *cómo* se evalúa la calidad de sitios o sedes web.
- c) Crear un protocolo propio, unificado y cuantificable, para la evaluación de sedes web –con parámetros e indicadores– a través de la implementación de una plantilla de evaluación de sitios o sedes web.

Para alcanzar estos objetivos será necesario, en primer lugar, presentar los criterios de diseño del corpus virtual y trilingüe que tenemos que compilar en el seno del proyecto TELL-ME, que se integrará por textos provenientes de la red Internet y que deberá cubrir los principales géneros y dominios de la Medicina

### 3. Diseño del corpus

Atendiendo a las necesidades del proyecto, a la Universidad de Málaga se encarga de compilar un corpus de texto con las siguientes características: *comparable*, pues se compondrá de textos originales, no traducidos; *trilingüe*, puesto que las lenguas que abarca el proyecto son el alemán, el español y el inglés; *especializado*, concretamente en el discurso científico; *textual*, ya que el corpus estará integrado por textos completos; y, *virtual*, debido a que todos los textos se descargarán directamente de la red Internet. Además, dado el dilatado número de países en los que el alemán, el español y el inglés son lenguas oficiales, el proyecto se encuentra delimitado diatópicamente a aquellas variedades habladas en Reino Unido, España y Alemania, por lo que los textos que se incluyan en el corpus también deberán estarlo.

Así, el corpus TELL-ME se integrará por tres subcorpus en función de la lengua (español, inglés y alemán); a su vez, cada subcorpus estará formado por cuatro componentes, según el género textual, a saber, artículos especializados procedentes de revistas especializadas, prospectos, artículos divulgativos procedentes de revistas de divulgación y artículos de divulgación procedentes de Wikipedia. Por último, en cada uno de los componentes mencionados, encontraremos 11 subcomponentes que contendrán textos que se enmarcan en un dominio específico de la medicina: primer contacto, urgencias, medicina interna, oftalmología, cirugía general, cirugía vascular, cirugía plástica, traumatología, otorrinolaringología, urología y ginecología.

### 4. Criterios de evaluación de los recursos electrónicos

Con el fin de asegurar la calidad de la muestra compilada, es fundamental abordar unos criterios de evaluación claros para los recursos que alberga la red Internet que tengan principalmente en cuenta *qué* y *cómo* se evalúa. Respecto a la cuestión de *qué* se evalúa en la red, podemos afirmar que son cinco<sup>4</sup> los documentos electrónicos susceptibles de ser evaluados, que se presentan a continuación atendiendo a una jerarquía conceptual descendente: en primer lugar, la *World Wide Web (WWW)*,

---

4 Estos cinco elementos se han extraído de los postulados de Codina (2000b) y pueden consultarse en: <http://www.ifla.org/VII/s13/pubs/isbd.htm>.

que es un sistema de información que permite explorar una red de redes de información digital basado en hipertexto (HTML). Precisamente, el hipertexto hace posible una estructura de consulta de una manera no lineal en la que, a través de un texto, se pueden acceder a otro documento que a su vez contiene más, y así de forma sucesiva. En segundo lugar, la *sede web* (también *sitio web* o *portal*), que es aquel que, dentro de la WWW, cuenta con una URL y alberga un conjunto de páginas web para facilitar su acceso y manejo. En tercer lugar, la *página web*, que es una de las páginas que integra un documento hipertextual de un sitio web. En cuarto lugar, la *página inicial* (también *principal* o *de inicio*) es la primera página que encuentra el usuario al acceder a un sitio web. Y, por último, la *URL*, que son las siglas de *Uniform Resource Locator*, es decir, estaríamos ante un, traducido literalmente, *localizador uniforme del recurso*, que puede ser definido como una cadena de caracteres con la cual se asigna una dirección única a cada uno de los recursos presentes en Internet. Una vez clarificados estos cinco conceptos, hemos de apuntar que, en lo a que a atañe a nuestro trabajo, diseñaremos una plantilla para la evaluación de sitios o sedes web principalmente pues, de una parte, dadas las dimensiones de Internet es prácticamente imposible el análisis de los cinco y, de otra, es precisamente en los sitios web a partir de donde se articula gran parte de la información contenida en ella.

Sobre la cuestión de *cómo* se evalúa, en toda evaluación, y la de recursos virtuales no podía ser menos, se precisan una serie de criterios, concretamente los denominados *parámetros e indicadores* de evaluación (cfr. Merlo Vega, 2003, pp. 101-110). Concretamente, los parámetros atienden a aquellos aspectos generales que son objeto de evaluación y que se desarrollan en indicadores concretos para el análisis de la misma. Son numerosas las monografías y artículos en los que se proponen parámetros e indicadores para determinar la calidad de los recursos digitales, muchos de ellos coincidentes a grandes rasgos, especialmente en lo que autoría y objetividad se refiere. Una de las propuestas más conocidas es la de Cooke (1999), quien considera que son diez —propósito, cobertura, autoridad y reputación, exactitud, actualidad y mantenimiento, accesibilidad, presentación y organización de la información, facilidad en el uso, comparación con otras fuentes y calidad— los parámetros que han de tenerse en cuenta a la hora de evaluar los recursos digitales. Sin embargo, a la mitad reduce esta cifra Auer (1998) para quien autoría, cobertura,



objetividad, exactitud y actualización de la información son criterios suficientes; en esta misma línea apunta la propuesta de Alexander y Tate (1999), quienes reflejan en su trabajo criterios semejantes, como los de autoría, exactitud, objetividad, actualidad y cobertura y destinatarios.

En el panorama nacional una de las propuestas más aceptada viene siendo la de Codina (2000a) quien, de los catorce<sup>5</sup> parámetros que contemplaba en un primer momento (cfr. Codina, 2000b), ocho pasarán a un segundo plano y considerará finalmente seis, con sus correspondientes indicadores, como aquellos esenciales a la hora de evaluar recursos electrónicos. De este modo, según este autor, la autoría, el contenido, el acceso, la ergonomía y la luminosidad son los seis parámetros que se han de considerar para el análisis de recursos digitales que, junto a sus respectivos indicadores, aparecen recogidos en la siguiente tabla (cfr. Tabla 1):

**Tabla 1. Parámetros e indicadores para la evaluación de recursos electrónicos formato (Codina, 2000a).**

Parámetros	Indicadores
Autoría	<i>Adecuación y solvencia del recurso.</i>
Contenido	<i>Valor y volumen de la información.</i>
Acceso	<i>Navegación y recuperación de la información.</i>
Ergonomía	<i>Comodidad y facilidad en el manejo.</i>
Luminosidad	<i>Presencia y calidad de nodos hipertextuales.</i>
Visibilidad	<i>Consideración del número de enlaces que recibe de otros recursos.</i>

Estas propuestas se han visto materializadas, en ocasiones, en plantillas de evaluación con el fin de ofrecer, en un formato más reducido y manejable, los criterios básicos para la evaluación de la información digital. Sin embargo, y a pesar de que parece no existir consenso en torno a cuáles deberían ser los criterios básicos para la evaluación de la información digital, sí es posible establecer algunos puntos en común.

5 Los catorce parámetros para la evaluación de recursos electrónicos que estableció Codina (2000b) fueron los de calidad y volumen de la información; autoría; legibilidad y ergonomía; navegación y representación de la información; recuperabilidad; interactividad y computabilidad; velocidad de descarga; servicios adicionales; luminosidad; calidad de los enlaces; actualización de los enlaces; descripción, selección y evaluación; visibilidad y autodescripción.



#### 4.1. Autoría, objetividad y actualización

La autoría, la objetividad y la actualización del contenido web parecen ser los tres pilares sobre los que se sustentan la mayoría de propuestas para el análisis de la calidad. Por lo que se refiere al *autor*, es esencial a la hora de determinar la fiabilidad de una sede web, y sus indicadores pueden ser el nombre de la persona o institución que la creó, así como sus credenciales y una dirección postal de contacto. En este sentido, la URL, con especial referencia al dominio, o la introducción del nombre en un motor de búsqueda —para comprobar cuántas veces aparece citado—, pueden corroborar si, efectivamente, la información sobre el autor es veraz o no.

A la hora de determinar el parámetro de la *objetividad*, entre los indicadores más comunes se encuentran la comprobación de elementos publicitarios, que puedan comprometer la objetividad del autor, así como el examen de si verdaderamente los objetivos que pretende cubrir la página son alcanzados o no, y si en ellos se vislumbra el tinte subjetivo del autor.

La consideración del parámetro de *actualización* es fundamental, pues nos dará cuenta de si la información contenida está desfasada o si, por el contrario, es actual. Entre los indicadores que han de tenerse en cuenta destacan la comprobación de la fecha de creación, así como de última actualización de la página. El silencio en torno a estos datos debe llevar al usuario a ser cauto en cuanto a la actualidad del contenido. No obstante, la comprobación del correcto funcionamiento de los distintos enlaces que contenga la página será síntoma de fiabilidad.

#### 4.2. Precisión

Otros autores cargan el peso a la hora de inclinar la balanza positivamente en la evaluación de un sitio web en la precisión, un criterio que, al igual que los tres expuestos más arriba, puede aparecer bien como parámetro, bien como indicador del parámetro *contenido*. Aunque suele ir acompañado de otros criterios de evaluación —normalmente, aunque no de forma exclusiva, de la *autoría*—, es considerado esencial para determinar la fiabilidad de la información albergada pues afecta específicamente a la calidad de la información contenida. De esta forma, la *precisión* persigue saber hasta qué punto se alcanzan los objetivos informativos.

### **4.3. Diseño**

El indicador que suele determinar el correcto diseño de una página es, principalmente, la adecuada estructuración de la información —que enlaza asimismo con la ausencia de errores— además de otros aspectos como la presencia de gráficos, vídeos o sonido así como los colores o tipo de fuente utilizada o el formato, menos relevantes pero que pueden favorecer o ralentizar el acceso. Todos estos factores que afectan al diseño deben aplicarse en aras de una mejor legibilidad y una menor pérdida de tiempo del usuario para localizar aquello que busca.

## **5. Protocolo unificado y cuantificable para la evaluación de sedes web**

Las reflexiones en torno a cuáles deberían ser los parámetros e indicadores imprescindibles son numerosas a la hora de evaluar un sitio o sede web. Desde la democratización informática, se han escrito ríos de tinta sobre estos conceptos y parece que una posición unánime, al menos a corto plazo, está lejos de ser alcanzada. Esto puede deberse, a nuestro entender, a tres factores principales que deben tenerse en cuenta: por un lado, las necesidades informativas del usuario (según el grado de especialidad, por ejemplo); por otro, el tiempo del que se disponga para realizar dicho análisis; y, finalmente, el volumen de datos objeto de análisis, ya que no es lo mismo visitar un único sitio web que un elevado número de ellos.

### **5.1. Parámetros e indicadores**

A tenor de lo anteriormente expuesto, y por razones de operatividad, para la plantilla de evaluación de los sitios web que servirá para garantizar la calidad de los textos que conformarán el corpus —que, a su vez, será la base para el desarrollo de herramientas de enseñanza-aprendizaje del proyecto TELL-ME—, nos centraremos en aquellos parámetros e indicadores que marquen con rapidez y precisión la fiabilidad del sitio web y no nos detendremos, por evidentes restricciones de tiempo y espacio, en otros (como por ejemplo el logotipo) que ralenticen y conviertan en tediosa la labor. Son numerosos los parámetros y aún más los

indicadores que pueden ser tenidos en cuenta a la hora de evaluar un sitio de Internet. Si el número de sedes que consultamos fuera reducido podríamos detenernos en el análisis de hasta incluso los 509 criterios que exponen investigadores como Wilkinson et al. (1997), cantidad que estimamos desorbitada; sin embargo, no es éste el caso, puesto que en trabajos de investigación como el que nos ocupa se requieren resultados fiables, rápidos y de calidad.

En este sentido proponemos tres parámetros básicos, como son *autoría*, *contenido* y, por último, *diseño* y *ergonomía*. En el caso del parámetro *autoría*, ha de ser personal o corporativo y no debe confundirse con el *webmaster*. Como indicadores, debe indicarse claramente el *nombre* del autor personal o corporativo y la *profesión*, *afiliación* u *organismo*, según el caso. Si se trata de un organismo o institución, debe recogerse de qué tipo, mientras que, si nos encontramos ante una persona física, la única vía que tenemos para identificar el grado de especialización del emisor es que éste indique su profesión o afiliación a un determinado organismo. Otro indicador fundamental en este parámetro a la hora de determinar la fiabilidad será la *URL*, dado que nos aporta una información esencial que puede llevarnos incluso a desechar o a almacenar la página antes de llegar a abrirla. De este modo, dominios del tipo *.edu*, *.gov*, *.int*, *.org* o *.prof* están reservados exclusivamente a instituciones u organizaciones intergubernamentales, a organizaciones en general y a profesionales (médicos, enfermeros, etc.), lo que les presupone un alto grado de fiabilidad y calidad; sin embargo, dominios como *.name*, destinados a personas individuales, o archivos como *~users*, *~members* o *~people*, nos invitan a ser extremadamente cautos con la información que contienen. En el caso concreto del proyecto TELL-ME, además, el dominio se perfila como esencial a la hora de aceptar un recurso electrónico, dado los límites diatópicos previamente establecidos, centrados en España, Alemania y Reino Unido. Así pues, serán de interés principalmente los dominios *.es*, para España; *.de*, para Alemania; o, *.uk*, para Reino Unido. Otros dominios indicativos de países serán descartados, ya que no son pertinentes para nuestra presente investigación. Nos referimos, por ejemplo, a páginas en español que contienen *.ar* (Argentina) o *.mx* (México); en alemán, *.ch* (Suiza); o en inglés, *.au* (Autralia), *.us* (Estados Unidos), o *.ie* (Irlanda).

Por lo que se refiere al parámetro *contenido*, serán de interés para compilar el corpus TELL-ME los indicadores como la *especialización*,

principalmente aquella media o alta (además, si se desea, podrá indicarse el *destinatario*); la *objetividad*, donde, de una parte, se eliminarán aquellos sitios web en los que se pueda ver cualquier tipo de sesgo como el ideológico, por ejemplo; de otra, el hecho de que la página cuente con referencias bibliográficas o se apoye en otras fuentes puede ser un indicio positivo; la *precisión*, pues se descartarán aquellos sitios web en los que se detecten errores y se valorarán aquellos que recojan enlaces a otros de interés; y la *actualización*, dado que no podrán incorporarse documentos con antigüedad de más de cinco años (2007) para asegurar la actualización de los contenidos (expresiones y vocabulario recientes). En caso de que no se indique la fecha de creación o de la última actualización, se comprobará el correcto funcionamiento de los enlaces, que puede servir de indicador de que el sitio web está siendo actualizado; por último, dado que el corpus es comparable, se atenderá a si la *lengua* en la que se ha redactado el documento es original o, en caso en que se trate de una traducción, se descartaría la información.

El último parámetro es aquel denominado *diseño y ergonomía*, cuyo indicador fundamental es una organización y estructuración lógica de la información, aspecto íntimamente relacionado con la especialización del autor, y que permitirá o dificultará el desenvolvimiento del internauta en la consulta de la información.

Por último, creemos importante recoger siempre la denominación del sitios web que analicemos, el *título*, aunque será un valor meramente informativo para la clasificación de la ficha generada tras la aplicación de la plantilla y no un parámetro ya que, en ocasiones, las sedes carecen de éste y, si cuentan con él, no siempre se ajusta al verdadero contenido de ella.

Una vez establecidos los parámetros e indicadores que nos permitirán evaluar la calidad de la información digital (sitios web) a partir de la cual se compilará el corpus TELL-ME, estamos en disposición de implementar una plantilla de evaluación.

## 5.2. Plantilla de evaluación

De este modo, la plantilla de evaluación para evaluar la calidad de los textos de sedes o sitios web que compondrán el corpus del proyecto TELL-ME contiene los parámetros e indicadores recogidos en el subapartado anterior (5.1.):

<b>TÍTULO:</b>	
<b>AUTORÍA</b>	
Nombre	
Profesión, afiliación, tipo de organismo	
URL	Diatópicas: <input type="checkbox"/> .es <input type="checkbox"/> .de <input type="checkbox"/> .uk Institucionales: <input type="checkbox"/> .com <input type="checkbox"/> .edu <input type="checkbox"/> .int <input type="checkbox"/> .org <input type="checkbox"/> Personales (~, %, users, members, people o name) <input type="checkbox"/>
Puntuación <sup>6</sup> 1 2 3 4 5	
<b>CONTENIDO</b>	
Tipo de información	Especialización baja <input type="checkbox"/> media <input type="checkbox"/> alta <input type="checkbox"/> (destinatario _____)
Objetividad	Sí <input type="checkbox"/> No <input type="checkbox"/>
Precisión	Sí <input type="checkbox"/> No <input type="checkbox"/> Errores detectados (ortotipográficos, terminológicos, etc...) Sí <input type="checkbox"/> No <input type="checkbox"/> Enlaces a otros sitios de interés Sí <input type="checkbox"/> No <input type="checkbox"/>
Actualización	Fecha de la última actualización _____ Fecha de creación _____ Funcionamiento de enlaces Sí <input type="checkbox"/> No <input type="checkbox"/>
Lengua/s	Original/es <input type="checkbox"/> Traducción/es <input type="checkbox"/>
Puntuación 1 2 3 4 5	
<b>DISEÑO Y ERGONOMÍA</b>	
Organización y estructuración lógica de la información	Sí <input type="checkbox"/> No <input type="checkbox"/>
Puntuación 1 2 3 4 5	
<i>Puntuación total:</i>	
12-15 puntos: <i>Fiabilidad alta</i> 9-11 puntos: <i>Fiabilidad media</i> 0-8 puntos: <i>Fiabilidad baja</i>	

**Figura 2. Parámetros e indicadores de la plantilla de evaluación puntuable.**

Para su implementación se ha recurrido Microsoft Excel (cfr. Figura 3), de modo que la plantilla se convierta en un elemento versátil y dinámico en la que el usuario con sólo hacer clic con el ratón señalará si un indicador se cumple o no o, en caso de que se cumpla, de qué tipo es; además,

<sup>6</sup> Valoración crítica de la sede o página según el parámetro objeto de análisis. La fiabilidad se puntuará objetivamente por el usuario del siguiente modo: 1= baja; 2= media-baja; 3= media; 4= media-alta y 5= alta.

le otorga un grado de fiabilidad al parámetro, con una puntuación individual entre 1 y 5, siendo 1 el valor más bajo y 5 el máximo. Esta cantidad procederá a sumarse automáticamente en la sección de *puntuación*.

	A	B	C
2	<b>TÍTULO:</b>		
3	<b>AUTORÍA</b>		
4	Nombre		
5	Profesión, afiliación, tipo de organismo		
6	URL	Diatópicas:	
7		Personales	.es .de .uk
8		Institucionales	
9	Puntuación 1 2 3 4 5		5
10	<b>CONTENIDO</b>		
11	Tipo de información	Especialización	
12		Destinatario	
13	Objetividad	Sí/No	
14			
15	Precisión	Sí/No	

Figura 3. Plantilla de evaluación puntuable implementada en Excel.

A cada parámetro se le asignan una serie de indicadores mínimos de campos rellenables que deben estar presentes en toda sede web evaluada puesto que, de lo contrario, se cuestionará el contenido de ésta y, por ende, su calidad. Además, en *puntuación total*, y tras haber valorado de 1 a 5 cada parámetro, se descartarán aquellas sedes cuya suma no alcance los 9 puntos; se considerarán con cautela aquellas de entre 9 y 11 puntos, con vistas a que puedan volver a ser consultadas y examinadas más adelante; y, finalmente, serán tenidas en cuenta para este proyecto las que hayan sumado de 12 puntos en adelante, en aras de la construcción de un corpus de calidad.

El umbral de puntuación ha sido fijado en función de un estudio piloto realizado con el objeto de comprobar la viabilidad de esta propuesta. En dicho estudio piloto se tomaron un total de 12 sedes web, a saber: tres revistas especializadas (como *BMJ*<sup>7</sup>, en inglés; *SEMES*,<sup>8</sup> en español; y, *Ärzte Zeitung*,<sup>9</sup> en alemán); tres revistas divulgativas (*Medline Plus*<sup>10</sup>, en inglés; *Consumer*<sup>11</sup>, en español y *Onmeda.de*<sup>12</sup>, en alemán); tres sedes especializadas en prospectos (*eMC*<sup>13</sup>, en inglés; *Vademecum*<sup>14</sup>, en español; y, *Diagnosia.de*<sup>15</sup>, en alemán); y, tres páginas personales<sup>16</sup>.

Efectivamente, se corroboró que las sedes de revistas especializadas y aquellas especializadas en prospectos alcanzaron, en las seis ocasiones, la máxima puntuación (15 puntos). Asimismo, todas las sedes divulgativas, emitidas por profesionales internacionales y nacionales, lograron alcanzar al menos los 12 puntos mínimos para su consideración (aquellas que no alcanzaron la cifra de 15 fue, principalmente, por la falta de objetividad, en los momentos en los que se vislumbraron fines eminentemente comerciales). Por último, de los sitios web personales manejados, superaron la cifra marcada aquellos que habían sido redactados por profesionales reconocidos del sector, como miembros de institutos de

---

7 <<http://www.bmj.com/>>.

8 <[http://www.semes.org/revista\\_EMERGENCIAS/](http://www.semes.org/revista_EMERGENCIAS/)>.

9 <<http://www.aerztezeitung.de/>>.

10 <<http://www.nlm.nih.gov/medlineplus/>>.

11 <<http://www.consumer.es/>>.

12 <<http://www.onmeda.de/>>.

13 <<http://www.medicines.org.uk/emc/>>.

14 <<http://www.vademecum.es/>>.

15 <<http://www.diagnosia.com/de/>>.

16 Se eligieron tres páginas personales, una por cada lengua de trabajo, de profesionales del sector de la medicina. No se han incluido explícitamente para preservar los datos personales.



investigación (sólo una de las sedes). Las dos sedes restantes, obtuvieron puntuaciones muy discretas (9 y 10 puntos), principalmente debido al silencio en torno a las fechas de actualización y creación, así como la falta de precisión y sus fines comerciales.

## 6. Conclusiones

En el caso específico de nuestra investigación, en un primer acercamiento, indudablemente, el resultado de los rastreos en la red, desde el punto de vista numérico, fue claramente abrumador; sin embargo, considerando que, hoy por hoy, se vuelca prácticamente toda la información en Internet sin pasar por ningún tipo de filtro de calidad, nuestra tarea primordial pasó de, limitarnos a la pura búsqueda de material, a llevar a cabo un análisis y evaluación del mismo para una posterior selección razonada con objeto de compilar el corpus TELL-ME, que servirá de base para la creación de herramientas en tres idiomas —inglés, español y alemán— para la enseñanza-aprendizaje de la comunicación médica. En este sentido, y como hemos apuntado previamente, la calidad es fundamental para la construcción de un corpus. Por consiguiente, es primordial saber discernir entre la información fiable presente en la red de aquella que no lo es. Para ello, hemos establecido unos criterios claros para la evaluación de recursos electrónicos que tienen en cuenta *qué* se evalúa, en función de una jerarquía conceptual descendente, a saber, la World Wide Web (WWW), la sede web (también sitio web o portal), la página web, la página inicial (también principal o de inicio) y la URL. Tras realizar un breve excursus por los principales parámetros e indicadores que permiten al usuario saber cómo se evalúa la calidad de sitios o sedes web, objeto de estudio de nuestro trabajo, hemos detectado que, en ocasiones, se manejan cifras desorbitadas que pueden llegar a ralentizar extremadamente el trabajo de evaluación de la información digital (como los 509 criterios que recogen investigadores como Wilkinson et al., 1997) así como una falta de consenso generalizada. Así las cosas, hemos creado un protocolo propio, unificado y cuantificable, para la evaluación de sedes web con únicamente tres parámetros —autoría, contenido, y diseño y ergonomía— acompañados de sus correspondientes indicadores. Todo ello se ha materializado a través de la creación de una plantilla puntuable de evaluación de sitios o sedes web (cfr. Figura

2). El umbral de puntuación ha sido fijado gracias a la realización de un estudio piloto donde se analizaron un total de 12 sedes web de revistas especializadas, revistas divulgativas, sedes especializadas en prospectos y páginas personales. Esta plantilla se ha implementado en formato Excel (cfr. Figura 3) para mantener un registro exhaustivo, sistemático y holístico (cf. Angelelli, 2009, p. 38), a la par que facilita la cuantificación automática y el almacenamiento de los resultados.

Asimismo, quisiéramos apuntar que hemos aplicado la plantilla creada antes de validar y de dar por fiable cualquier sitio web destinado a formar parte del corpus trilingüe de medicina compilado en el seno del proyecto TELL-ME. Una vez que el sitio web ha superado la puntuación mínima requerida (12 puntos) se procede a la extracción automática de los textos que lo integran con el programa Bootcat<sup>17</sup>. El resultado ha sido un corpus –el corpus TELL-ME–, representativo, fiable y de calidad, gracias a la plantilla de evaluación de sedes web, y que actualmente se encuentra integrado por 149.559 types (palabras) y 10.823.893 tokens (conurrencias) en inglés; 121.244 types y 4.711.886 tokens en español; y 172.200 types y 8.163.489 tokens en alemán, a partir del cual se han desarrollado todas las herramientas de enseñanza-aprendizaje de la plataforma TELL-ME.

Por último, cabe destacar que la plantilla diseñada puede ser utilizada, asimismo, para la evaluación de cualquier recurso científico en red en alemán, inglés y español.

## Referencias bibliográficas

- ALEXANDER, J.E. y TATE, M. (1999). *Web wisdom: how to evaluate and create information quality on the Internet*. Mahwah, Lawrence Erlbaum Associates.
- ANGELELLI, C. V. (2009). Using a rubric to assess translation ability: defining the construct. En C. Angelelli y H. Jacobson (eds.), *Testing and Assessment in Translation and Interpreting Studies* (pp. 13-47). Ámsterdam: John Benjamins.
- ASTORELLI, C. (2012). Encontrar trabajo fuera de España. *Eroski consumer*, 12, (3), 2-10.
- AUER, N.J. (1998). Bibliography on Evaluating Internet Resources. *Emergency Librarian*, 25, (5), 23-24.
- AUSTERMÜHL, F. (2001). *Electronic Tools for Translators*. Manchester, St. Jerome.

17 El programa para la extracción de corpus a partir de sedes web, Bootcat, puede descargarse en la siguiente dirección: <http://bootcat.sslmit.unibo.it>. Para una descripción de la metodología para la extracción de corpus de forma semiautomática con Bootcat consúltese Seghiri et al. (2013).

- CODINA, L. (2000a). Parámetros e indicadores de calidad para la evaluación de recursos digitales. En A. Buzón (ed.), *La Gestión del Conocimiento: retos y soluciones de los profesionales de la información* (pp. 135-144). Bilbao: Universidad del País Vasco.
- CODINA, L. (2000b). Evaluación. *Revista Española de Documentación Científica*, 23, (1), 9-44.
- COOKE, A. (1999). *Neal-Schuman authoritative guide to evaluating information on the Internet*. Nueva York, Neal-Schuman.
- MERLOVEGA, J.A. (2003). La evaluación de la calidad de la información web: aportaciones teóricas y experiencias prácticas. En A. Felipe ZAPICO (ed.), *Recursos informativos: creación, descripción y evaluación* (pp. 101-110). Mérida: Junta de Extremadura.
- RIVERA, R. (2013). España quiere que el turismo sanitario se convierta en un negocio. *Teinteresa.es*, 10 de enero, 3-12.
- SEGHIRI, M. (2011). Metodología protocolizada de compilación de un corpus de seguros de viajes: aspectos de diseño y representatividad. *Revista de lingüística teórica y aplicada*, 49, (2), 13-30.
- SEGHIRI, M., CORPAS-PASTOR, G. y GUTIÉRREZ-FLORIDO, R. (2013). Using semi-automatic compiled corpora for medical terminology and vocabulary building in the healthcare domain. En R. Temmerman y G. Aguado de Cea (eds.), *Optimizing understanding in multilingual hospital encounters* (pp. 5-18). París: Université of Paris XIII.
- WILKINSON, G.L., BENNETT, L.T. y OLIVER, K.M. (1997). *Evaluating the Quality of Internet Information Sources*. Calgary, ED-MEDIA/ED-TELECOM.